Gene, 81 (1989) 45–54
Elsevier

GENE 03100

Nucleotide sequence and analysis of the lethal factor gene (*lef*) from *Bacillus anthracis*

(Recombinant DNA; exotoxin; Gram-positive; signal peptide; leader; secretion; *B. subtilis*; protein transport)

Thomas S. Bragg and Donald L. Robertson

Graduate Section of Biochemistry, Department of Chemistry, Brigham Young University, Provo, UT 84602 (U.S.A.)

Received by R.E. Yasbin: 7 February 1989
Revised: 27 March 1989
Accepted: 3 April 1989

SUMMARY

The nucleotide sequence of the *Bacillus anthracis* lethal factor (LF) gene (*lef*) has been determined. LF is part of the tripartite protein exotoxin of *B. anthracis* along with protective antigen (PA) and edema factor (EF). The apparent ATG start codon, which is located immediately upstream from codons which specify the first 16 amino acids (aa) of the mature secreted LF, is preceded by an AAAGGAG sequence, which is its probable ribosome-binding site. This ATG codon begins a continuous 2427-bp open reading frame which encodes the 809-aa LF-precursor protein with an M_r of 93798. The mature secreted protein (776 aa; M_r 90237) was preceded by a 33-aa signal peptide which has characteristics in common with leader peptides for other secreted proteins of the *Bacillus* species. The codon usage of the LF gene reflects its high (70%) A + T content. The N-terminus of LF (first 300 aa) shared extensive homology with the N-terminus of the anthrax EF protein. Since LF and EF each bind PA at the same site, these homologous regions probably represent their common PA-binding domains.

INTRODUCTION

Bacillus anthracis, which causes anthrax, infects many mammalian species, including humans. The virulence of anthrax bacilli is due to the production of a poly-D-glutamic acid capsule and a three-component protein exotoxin (Leppla et al., 1985). The toxin proteins have been purified and consist of PA, EF and LF (Beall et al., 1982; Ezzell et al., 1984; Leppla et al., 1985). The lethal toxin, which contains PA and LF, causes death in rats, guinea pigs and mice (Beall et al., 1982; Little and Knudson, 1986). The edema toxin, which is composed of PA and EF, produces a localized edema in the skin of guinea pigs and rabbits (Stanley et al., 1960; Thorne et al., 1960). Each of the anthrax toxin genes has been cloned (Robertson and Leppla, 1986; Tippetts and Robertson, 1988; Vodkin and Leppla, 1983; Mock

Correspondence to: Dr. D.L. Robertson, 659 WIDB, Brigham Young University, Provo, UT 84602 (U.S.A.) Tel. (801) 378-7018; Fax. 801-378-5474.

Abbreviations: aa, amino acid(s); *B.*, *Bacillus*; bp, base pair(s); *cya*, gene coding for EF; EF, edema factor (adenylate cyclase); kb, kilobase(s) or 1000 bp; *lef*, gene coding for LF; LF, lethal factor; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; ORF, open reading frame; *pag*, gene coding for PA; PA, protective antigen; RBS, r.b.s., ribosome binding site(s); ss, single strand(ed); *tsp*, transcription start point(s); USAMRIID, United States Army Medical Research Institute of Infectious Diseases (Frederick, MD).

et al., 1988) and the nt sequences for the PA (*pag*) and EF (*cya*) genes have recently been reported (Welkos et al., 1988; Robertson et al., 1988).

EF is a calmodulin-dependent adenylate cyclase (Leppla, 1982, 1984; Leppla et al., 1985) which may be involved in the formation of the edematous lesions in cutaneous anthrax. EF shares extensive aa and nt sequence homology with the *Bordetella pertussis* calmodulin-dependent adenylate cyclase (Robertson et al., 1988), but is not related to the *Escherichia coli* or yeast adenylate cyclases. The lethal effects of LF are well documented, but the mode of action is not known, although cell disruption apparently occurs (Friedlander, 1986). PA apparently has no enzymatic activity, but is required for toxin uptake.

In order for the toxin proteins to enter a mammalian cell, PA must first bind to a cellular receptor, the identity of which has not yet been determined. Once bound, PA is activated by proteolytic cleavage which results in the release of a 20-kDa N-terminal polypeptide. After cleavage and removal of this polypeptide, a domain(s) which binds LF and EF is exposed. The binding of LF or EF to the modified PA is competitive, suggesting that these proteins bind PA at the same site (S.H. Leppla, personal communication). After LF or EF is bound to PA, the entire toxin complex apparently enters the cell by endocytosis (Leppla, 1984; Leppla et al., 1985; Friedlander, 1986; S.H. Leppla and A.M. Friedlander, personal communications). Friedlander (1986) has shown that for LF, at least, an acid-dependent step is utilized. Since LF and EF appear to compete with each other for binding to PA, it is anticipated that LF and EF should share at least some aa sequence homology.

The goal of our studies on the anthrax toxin genes is to use our recombinant clones to develop a more effective human anthrax vaccine. We should also be able to better understand the role of EF and LF in the pathogenesis of *B. anthracis*. In addition, our studies should help elucidate how EF and LF interact with PA and how these toxins enter the cell. In this communication, we describe the complete nt sequence and the deduced aa sequence for the LF gene (*lef*). We also show that LF and EF share a highly conserved N-terminus, which is probably required to bind PA prior to cellular uptake.

MATERIALS AND METHODS

(a) Reagents and enzymes

DNA restriction and modifying enzymes were obtained from Bethesda Research Laboratories (Gaithersburg, MD). For sequencing deoxyribonucleoside and dideoxyribonucleoside triphosphates, as well as the modified T7 DNA polymerase (Sequenase), were purchased from U.S. Biochemical Corp. (Cleveland, OH). Deoxyribonucleoside [α-^{32}P]triphosphates (800 Curies/mmole) were obtained from New England Nuclear (Boston, MA).

(b) Nucleotide sequence analysis

Nucleotide sequencing was performed using the Sequenase DNA sequencing kit from U.S. Biochemical with minor modifications. The initial sequencing reaction used a synthetic mixed oligo primer specific for the N-terminus of LF and an ss plasmid, (pLF74, which contained the 2.1 kb *Pst*I DNA fragment from pLF7) containing part of the *lef* gene which apparently includes the start codon and *tsp* (Robertson and Leppla, 1986). However, in order to sequence the entire *lef* gene, synthetic oligo primers 18–22 nt long were used to obtain the rest of the sequence, using plasmids which contained the 4.9-kb *Bam*HI-*Sst*I DNA fragment from pLF7 (see Fig. 1A). Both strands were sequenced. The sequencing oligos were prepared using β-cyanoethyl phosphoramidites on an Applied Biosystems (Foster City, CA) DNA synthesizer. The sequencing reaction products were separated in denaturing 8% polyacrylamide gels. After autoradiography, the nt sequences were recorded using a Beckman GelMate 1000 Sonic Digitizer. Individual sequences were merged and analyzed using the MicroGenie Sequence Software from Beckman Instruments (Palo Alto, CA).

RESULTS AND DISCUSSION

(a) Nucleotide sequence analysis of the lethal factor gene (*lef*)

Figure 1A shows a recombinant plasmid containing the entire *B. anthracis lef* gene (Robertson

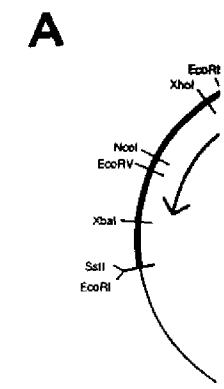


Fig. 1. Recombinant pla
quencing strategy. (A) R
from pXO1 (Kaspar and
solid line, and from pT2
The arrow inside the ci
translation and shows t
Several restriction sites
of the *B. anthracis lef* g
using the Sanger et al.
techniques. Each of the
sequence determinations
of the restriction sites d
shown and have been cor

and Leppla, 1986)
plasmid pXO1 (K
Using a synthetic oli
LF (Robertson and
(using primer elong
about 600 bp upstr
transcription proce
ognition site (see
authors). Our initial
that the nt sequence
site of this LF-specifi
the first 16 aa of LI
the aa sequence at t
USAMRIID, perso

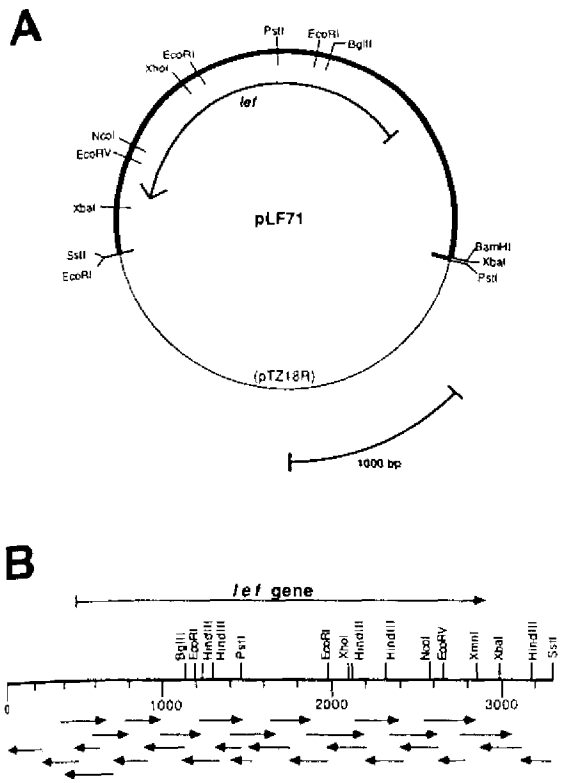


Fig. 1. Recombinant plasmid containing the *lef* gene and sequencing strategy. (**A**) Restriction map of pLF71. DNA which is from pXO1 (Kaspar and Robertson, 1987) is shown as the heavy solid line, and from pTZ18R (Mead et al., 1986) as a thin line. The arrow inside the circle depicts the starting point for LF translation and shows the direction for *lef* gene transcription. Several restriction sites are also shown. (**B**) Sequencing strategy of the *B. anthracis lef* gene. The nt sequence was determined using the Sanger et al. (1977) dideoxynucleotide termination techniques. Each of the short arrows represents independent nt sequence determinations using different synthetic oligos. Some of the restriction sites deduced from the nt sequence are also shown and have been confirmed by restriction enzyme digestion.

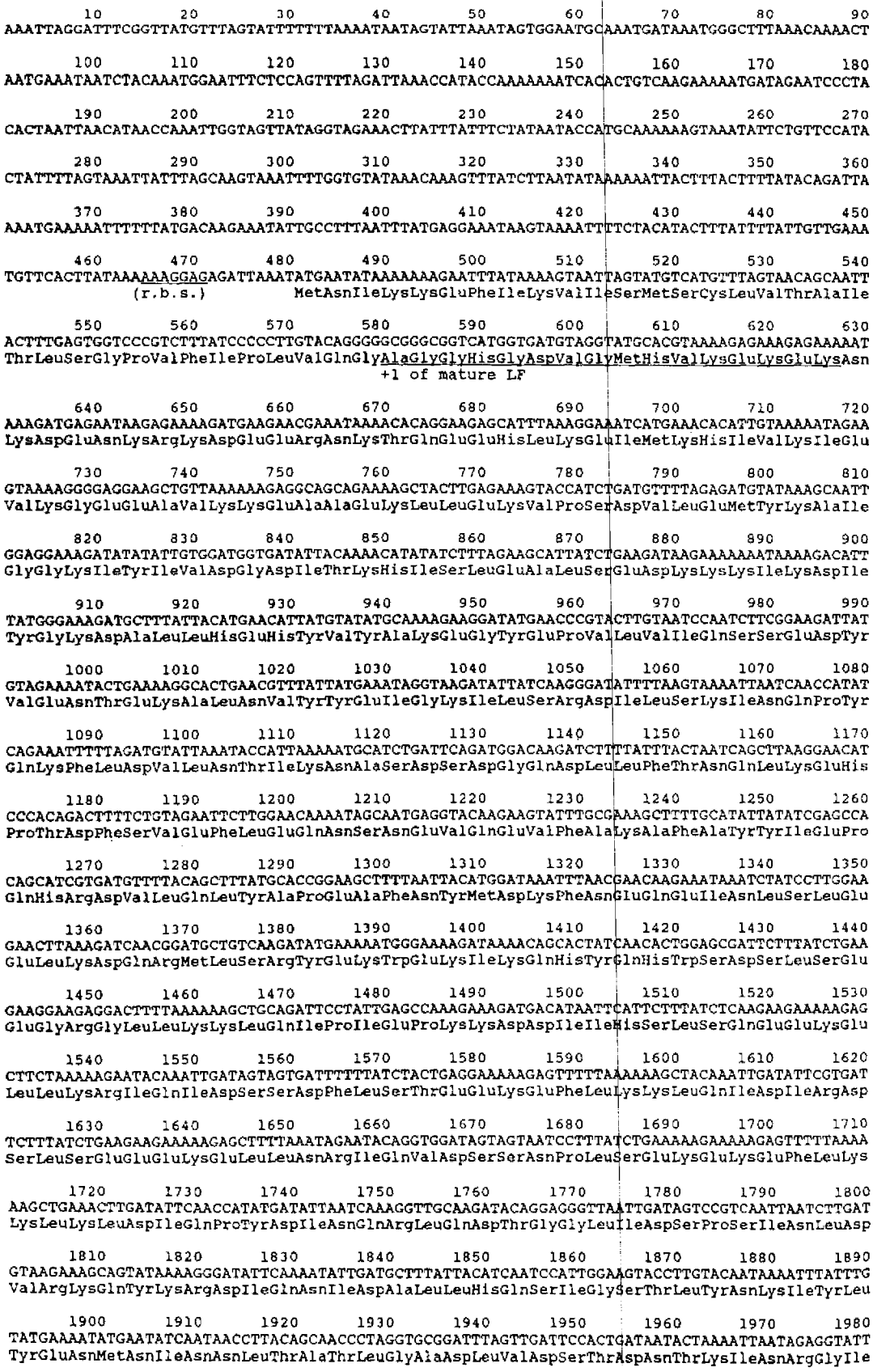
and Leppla, 1986) isolated from the 176-kb toxin plasmid pXO1 (Kaspar and Robertson, 1987). Using a synthetic oligo specific for the N-terminus of LF (Robertson and Leppla, 1986), we identified (using primer elongation) the start of the *lef* gene about 600 bp upstream from the *Eco*RI site with transcription proceeding away from the *Bam*HI recognition site (see Fig. 1A; unpublished data of authors). Our initial nt sequencing reaction showed that the nt sequence downstream from the binding site of this LF-specific oligo contained the codons for the first 16 aa of LF (Fig. 2), completely matching the aa sequence at the LF N-terminus (J. Schmidt, USAMRIID, personal communication).

We sequenced almost 3300 bp, including the 5′- and 3′-noncoding flanking regions of the *lef* gene. Fig. 1B shows the sequencing strategy for determination of the *lef* gene sequence using the dideoxynucleotide termination technique described by Sanger et al. (1977). Synthetic oligos were used to prime DNA synthesis along the DNA in both directions (see MATERIALS AND METHODS, section **b**).

(b) Translation and transcription regulatory regions

Figure 2 shows the complete nt sequence of the *lef* structural gene with its flanking regions. There was a single long ORF, which encoded an 809-aa protein. The aa sequence for the first 16 aa residues of mature LF has been determined (J. Schmidt, personal communication) and is underlined in Fig. 2 (nt 580–627). The first ATG codon (nt 481) upstream from the codons, which specified the start of mature LF, was preceded by its probable RBS (AAAGGAG), located at nt positions 465–471. If this entire RBS sequence base-paired with the ribosome, the calculated free energy for this interaction would be −18.8 kcal/mole (−78.7 kJ/mole) (Tinoco et al., 1973), which is similar to RBS for genes from other Gram-positive bacteria (Schwarz et al., 1988). This nt sequence was identical to the probable RBS for the *pag* gene (Welkos et al., 1988) and close to the sequence for the *cya* gene (AAAGGAGGT) (Robertson et al., 1988). A consensus RBS for *B. anthracis* is not known, but the *cya*, *lef* and *pag* RBS sequences were complementary to the 3′-end of the 16S rRNA of the Gram-positive *B. subtilis* (Moran et al., 1982; Band and Henner, 1984). Since LF translation probably begins at the ATG codon ten nt downstream from this sequence, the LF-precursor would contain a 33-aa signal peptide, which is then cleaved to generate mature LF in *B. anthracis*. Proteolytic cleavage may not be required for enzyme activity, however, since LF isolated for *E. coli*, which was intracellular and probably not cleaved, is biochemically active (Robertson and Leppla, 1986).

The *lef* gene RBS appeared to function well in *B. anthracis*, *B. subtilis* (unpublished data of authors) and even in *E. coli* (Robertson and Leppla, 1986). We have not yet identified the position or sequence of either the *pag*, *lef* or *cya* promoters. Therefore, until S1 mapping or primer elongation experiments



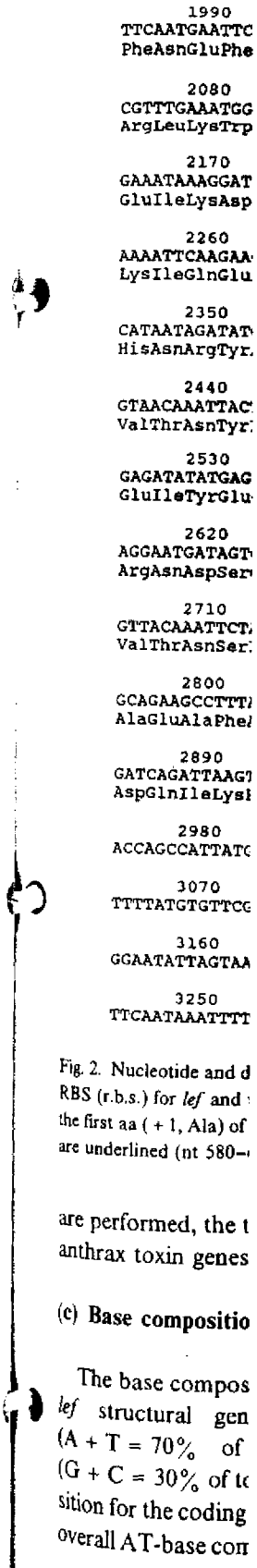


Fig. 2. Nucleotide and d
RBS (r.b.s.) for *lef* and
the first aa (+ 1, Ala) of
are underlined (nt 580–

are performed, the t
anthrax toxin genes

(c) **Base compositio**

The base compos
lef structural gen
(A + T = 70% of
(G + C = 30% of t
sition for the coding
overall AT-base com

Fig. 2. Nucleotide and deduced aa sequence for the *lef* structural gene with its 5′- and 3′-noncoding flanking regions. The presumptive RBS (r.b.s.) for *lef* and the probable start codon are shown. The 33-aa signal peptide which starts the LF ORF at nt 481, as well as the first aa (+1, Ala) of mature LF (nt 580) are also shown. The first 16 aa of mature LF, as determined by J. Schmidt (USAMRIID), are underlined (nt 580–627).

are performed, the transcription start sites for these anthrax toxin genes will remain unknown.

(c) Base composition and codon usage

The base composition of the coding strand of the *lef* structural gene was: A = 41%, T = 29% (A + T = 70% of total), G = 18%, C = 12% (G + C = 30% of total). The 70% AT base composition for the coding strand is slightly higher than the overall AT-base composition for pXO1 and genomic DNA, which are about 69% (Kaspar and Robertson, 1987). The 5′-noncoding region immediately upstream from the *lef* gene has a higher A + T content (78%), which seems to be characteristic of the regulatory regions for genes of bacilli and related bacteria (Moran et al., 1982). For example, the *pag* structural gene has an A + T content of 67% but contains a higher (75%) A + T base composition for its upstream regulatory sequences (Welkos et al., 1988).

The codon usage for the entire LF-precursor protein is shown in Table I. There is a preference for

codons which contain an A or T in the third position, which likely reflects the high A + T content of the gene. For example, codons for aa which have six codons (e.g., Leu, Ser and Arg), use the triplet combinations which have the higher A + T contents. Similar codon usage was observed for the *pag* and *cya* genes (Robertson et al., 1988; Welkos et al. 1988). Overall codon usage for *B. anthracis* is not known, but Shields and Sharp (1987) showed that highly expressed genes from several different unicellular organisms use codons for which the most abundant tRNAs are available.

(d) Amino acid sequence of the LF protein

Figure 2 also includes the deduced aa sequence for the full-length LF-precursor (809 aa) with an M_r of 93 798. Since the aa sequence of mature LF actually begins at aa position 34 of the LF-precursor (at the Ala residue marked +1; see RESULTS AND DISCUSSION, section **b**), the 33-aa leader peptide preceding this position must be removed during secretion. This signal peptide conforms to known *Bacillus* leader sequences in that it started with charged (mostly positive) and hydrophilic residues (aa 1–9), followed by a central core of hydrophobic aa (aa residues 10–22) and then several hydrophilic residues (aa 23–33) prior to the start of the mature protein. Proteolytic cleavage apparently occurs at a Gly-Ala peptide bond consistent with signal processing after an Ala or Gly in *Bacillus* spp. (Pugsley and Schwartz, 1985; MacKay et al., 1986; O'Neill et al., 1986). Fig. 3 shows the aa sequences near the ends of the EF, PA and LF signal peptides. Similar

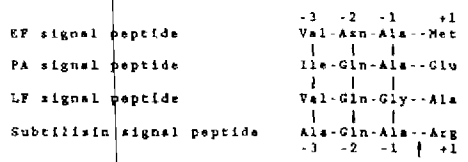


Fig. 3. Signal peptide analysis. The last three aa at the C-terminus of the deduced EF, PA and LF signal peptides are shown. Also included are the aa at the C-terminus of the *B. subtilis* subtilisin signal peptide. The numbers indicate the location of aa residues relative to the deduced signal peptide cleavage site (↑). Negative numbers (−1, −2 and −3) indicate their upstream positions, whereas the positive number (+1) indicates the downstream location.

TABLE 1

Codon utilization in the *lef* gene of *Bacillus anthracis*

Codon	aa	No.[a]	Codon	aa	No.[a]	Codon	aa	No.[a]	Codon	aa	No.[a]
TTT	Phe	21	TCT	Ser	19	TAT	Tyr	32	TGT	Cys	1
TTC[b]	Phe	8	TCC	Ser	6	TAC[b]	Tyr	3	TGC	Cys	0
TTA	Leu	43	TCA	Ser	8	TAA[c]	End[c]	0	TGA[c]	End[c]	0
TTG	Leu	8	TCG	Ser	2	TAG[c]	End[c]	0	TGG	Trp	5
CTT	Leu	16	CCT	Pro	5	CAT	His	16	CGT[b]	Arg	5
CTC	Leu	2	CCC	Pro	4	CAC	His	5	CGC[b]	Arg	0
CTA	Leu	6	CCA[b]	Pro	9	CAA	Gln	28	CGA	Arg	2
CTG[b]	Leu	5	CCG[b]	Pro	3	CAG[b]	Gln	13	CGG	Arg	1
ATT	Ile	41	ACT[b]	Thr	10	AAT	Asn	49	AGT	Ser	16
ATC[b]	Ile	7	ACC[b]	Thr	4	AAC[b]	Asn	8	AGC	Ser	2
ATA	Ile	26	ACA	Thr	13	AAA[b]	Lys	63	AGA	Arg	12
ATG	Met	10	ACG	Thr	1	AAG	Lys	23	AGG	Arg	7
GTT[b]	Val	12	GCT[b]	Ala	11	GAT	Asp	51	GGT[b]	Gly	14
GTC	Val	1	GCC	Ala	1	GAC	Asp	4	GGC[b]	Gly	1
GTA[b]	Val	22	GCA[b]	Ala	12	GAA[b]	Glu	59	GGA	Gly	14
GTG	Val	5	GCG[b]	Ala	5	GAG	Glu	20	GGG	Gly	8

[a] Number of codons in the entire *cya* coding region (see Fig. 2).

[b] Major *E. coli* tRNA species (Ikemura, 1981).

[c] Stop codons.

aa sequences are
peptides (Pugsley
quence for the en
(Wang et al., 1988
aa residues at the e
are probably requ
nition and cleavag

Mature LF, sta
sponding to nt 58
larger than the pr
83 kDa (Leppla et
of LF by Quinn et
electrophoretic mc
least very close t
89 kDa (Leppla et
We conclude, there
slightly larger than
aa sequence and s
content for LF is
derived values dete
LF (J. Schmidt, US
tion). It is also int
anthrax toxin prot
contains a single C
removed during se
for LF, which is cl
mined value of 5.8
larger number of ac
residues (110 aa) ir

Initially, we wer
size of LF, based o
previously reported
when we started t
observed than wh
compared to itself,
repeated regions lo
(see Fig. 4). We al
quence contains n
correspond to the
shown). It should b
repeated domains fc
cations, but repeats

Therefore, in ord
which we used for
during the process
specific DNA isolat
viously shown to en
and Leppla, 1986)]
(see MATERIALS AN

aa sequences are present for other *Bacillus* signal peptides (Pugsley and Schwartz, 1985). The aa sequence for the end of the subtilisin signal peptide (Wang et al., 1988) is also included in Fig. 3. Similar aa residues at the end of these *Bacillus* signal peptides are probably required for signal peptidase recognition and cleavage.

Mature LF, starting with aa position 34 (corresponding to nt 580), has an M_r of 90 237 which is larger than the previously reported value of about 83 kDa (Leppla et al., 1985). However, recent sizing of LF by Quinn et al. (1988) showed that LF has an electrophoretic mobility slightly slower than, or at least very close to, that for EF, which is about 89 kDa (Leppla et al., 1985; Robertson et al., 1988). We conclude, therefore, that LF should have an M_r slightly larger than EF, consistent with our deduced aa sequence and size calculations. The deduced aa content for LF is also close to the experimentally derived values determined from an acid hydrolysis of LF (J. Schmidt, USAMRIID, personal communication). It is also interesting that none of the mature anthrax toxin proteins contains Cys, although LF contains a single Cys in the signal peptide, which is removed during secretion. The calculated pI (6.01) for LF, which is close to the experimentally determined value of 5.8 (Quinn et al., 1988), reflected the larger number of acidic (133 aa), compared to basic, residues (110 aa) in the mature secreted protein.

Initially, we were concerned when the deduced size of LF, based on its nt sequence, was larger than previously reported (Leppla et al., 1985). In addition, when we started to analyze the LF sequence, we observed than when the aa sequence of LF was compared to itself, LF possessed several internal repeated regions located between aa 300 and 420 (see Fig. 4). We also observed that the *lef* nt sequence contains nt repeats in the regions which correspond to the repeated aa domains (data not shown). It should be emphasized, however, that the repeated domains for LF and *lef* are not exact duplications, but repeats possessing 80–90% homology.

Therefore, in order to be certain that the DNA which we used for nt sequencing was not altered during the process of cloning, we analyzed LF-specific DNA isolated from pLF7 [which was previously shown to encode functional LF (Robertson and Leppla, 1986)], pLF71 (Fig. 1A) and pLF74 (see MATERIALS AND METHODS, section **b**) (which were used for nt sequencing), and pXO1. We digested each of these DNAs with *Eco*RI or *Hin*dIII and then blotted the electrophoretically-separated DNA onto nitrocellulose. After hybridization with a LF-specific probe, we observed that each analyzed DNA contained identically sized DNA fragments containing the repeated regions for each enzyme. These results clearly show that our cloned DNAs, which produce active LF and which were used for nt sequencing, had been faithfully propagated in *E. coli* and are identical to the corresponding region in pXO1. Consequently, we feel that the nt sequence we determined for the region of *lef* which contains the internal repeats is correct and that *lef* contains bona fide repeats. The function, if any, of these repeated regions in LF is not known.

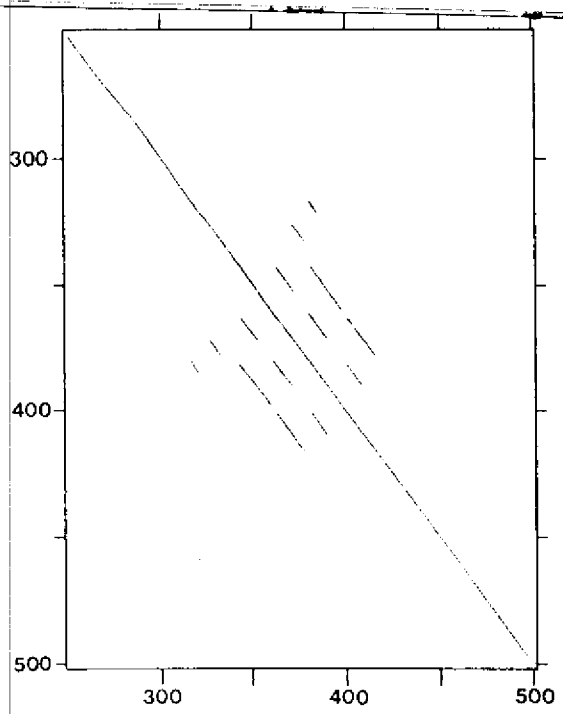


Fig. 4. Internal repeated regions for the aa sequence of LF. The entire LF aa sequence was compared to itself and found to possess several duplicated regions clustered between aa 300 and 420 (located between nt 1380 and 1740 of Fig. 2). It was also observed, although the data is not shown, that a similar set of repeats were present in the nt sequence of *lef* in the region which corresponds to these aa repeats. The numbers on each axis show the aa position relative to the entire aa sequence of the LF-precursor protein.

aa	No.[a]
Cys	1
Cys	0
End[c]	0
Trp	5
Arg	5
Arg	0
Arg	2
Arg	1
Ser	16
Ser	2
Arg	12
Arg	7
Gly	14
Gly	1
Gly	14
Gly	8

(e) Relationship of *Bacillus anthracis lef* and LF to other known genes and proteins

There is no detectable homology between the *B. anthracis lef* gene, or its deduced aa sequence, with any other known gene or protein from the current GenBank and NBRF databases (March 1988). However, we have observed that the N-terminus of LF is homologous to the corresponding N-terminal domain of EF. The regions of homology between these proteins are shown in Fig. 5. It is probable that these homologous aa domains are required to bind PA prior to cellular uptake. We also determined the hydropathic profiles for the LF- and EF-precursor proteins (Robertson et al., 1988), which are shown in Fig. 6. If these homologous domains are required for binding PA, then it would be anticipated that LF and EF should have similar hydropathic profiles. Our analysis indicated that the conserved domains of EF and LF, which are mostly hydrophilic, would probably be located on the surface where they could interact with PA.

Since the N-termini of LF and EF probably bind PA, it is presumed that the catalytic domains of LF and EF must reside within their respective C-terminal regions. For example, we have recently shown that the C-terminus of EF (aa 300–800) contains its ATP-binding site and is the region which is homologous with the *B. pertussis* calmodulin-dependent adenylate cyclase (Glaser et al., 1988; Robertson, 1988; Robertson et al., 1988). Fig. 7 shows the important structural domains of LF and EF, which are localized in their N-termini. The C-terminus of EF contains its catalytic domain, but while it is likely that the C-terminus of LF also contains its biochemical activity, we do not know what this activity might be.

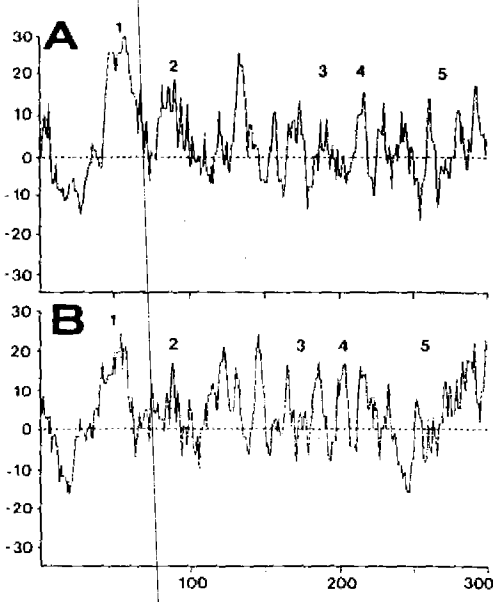


Fig. 6. Hydropathic analysis of the C-terminal domains of the (**A**) LF- and (**B**) EF-precursor proteins. Using the algorithm of Hopp and Woods (1981), the proposed hydropathic values were determined. The hydrophilic residues are positive and the actual calculated numbers ar multiplied by 10. (The signal peptides start with positive values (hydrophilic residues), and then the hydrophobic regions which have negative values.) The conserved regions between LF and EF (see Fig. 5) are numbered and have similar hydropathic characteristics.

```
LF 1   MNIKKEFIKVISMSCLVTAITLSGPVFIPLVQGAGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKHIVKIEVKGEEAVKKEAAEKLLEKV
EF 1   MT RNKFIPN KFSIISFSV L   LF AIS SSQAI EVNA MNEHYTESDIKRNHKTEKNKTEKEKFKDSINNLVKTEFTNETLDKIQQTQDLLKKI
                                                  1

100    PSDVLEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDALLHEHYVYAKEGYEPVLVIQSSEDYVENTEKALNVYYEIGKILSRDILSKINQ
 90    PKDVLEIYSELGGEIYFTDIDLVEHKELQDLSEEEKNSMNSRGEKVPFASRFVFEKKRETPKLIINIK DYAINSEQSKEVYYETGKGISLDIISKDKS
       2                                                                                                3

199    PYQKFLDVLNTIKNASDSDGQDLLFTNQLKE    HPTDFSVEFLEQNSNEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEINLSLEELK
188    LDPEFINLIKSISD  DSDSSDLLFSQKFKEKLELNNKSIDINFIKENLTEFQHAFSLAFSYYEAPDHRTVLELVAPDMFEYMNKLEKGGFEKISESLK
                  4                                                                 5

294    DQRMLSRYEKWEKIKQHYQHWSDSLSEEGRGLLKKLQIPIEPKKDDIIHSLSQEEKELLKRIQIDSSDPLSTEEKEFLKKLQ IDIRDSLS EEEKEL
285    KE GVEK DRIDVLKGEKALKASGLVPEHADAFKK  IARELNTYILFRPVNKLATNLIKSGVATKGLNFHGKSSDWGPVAGYIPFDQDLSKKHGQQLA
                                           a                                 b

390    LNRIQVDSSNPLSEKEKEFLKKLKLDIQPYDINQRLQDTGGLIDSPS
380    VEKGNLENKKSITEHEGE IGKIPLKLDHLRIEE LKENGIILKGKK
```

Fig. 5. Amino acid homology comparison between LF and EF. The aa sequences for the LF- and EF-precursor proteins were compared. Five regions of aa homology were observed (domains **1–5**). The arrows indicate the first aa of the mature, secreted proteins. Also shown are the probable calmodulin-binding site (domain **a**) and ATP-binding site (domain **b**) of EF, both of which are conserved in the *B. pertussis* calmodulin-dependent adenylate cyclase (Robertson, 1988; Robertson et al., 1988). Short vertical lines connect identical aa; colons connect similar aa.

Fig. 7. Structural domai
the first 300 aa of these
cyclase (Robertson, 198
probably occupies the c

(f) Conclusions and

We have cloned ar
gene, which encodes
protein exotoxin of *B*
has features in comi
coding (*pag*) and El
similar RBS and lo
apparently cleaved
et al., 1988; Welkos
each of these toxin ge
none of the mature pr
of mature LF, deduc
90 327 Da, which is cl
mined values (Lepp
1988).

We now know the
B. anthracis cya, lef a
information, we shou
sion vectors for prod
which can be used fc
well as studying the bi
toxin proteins. In add
experiments are in pr
required for activity,
duction of a safer rec
cine.

It is also of interest
by which EF and LF
previously, PA binds
is cleaved prior to bin
ments which are desig
300-aa N-terminus of
to transport heterologo
PA as the transporting
laboratory. The availa
mechanism to introduc

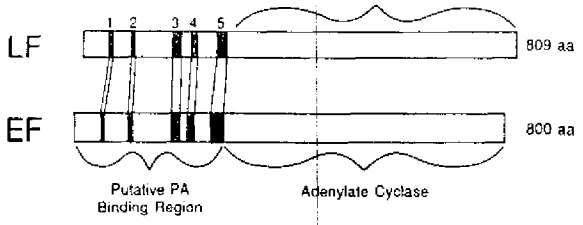


Fig. 7. Structural domains of LF and EF. The putative PA binding domains for LF and EF (see Fig. 5) are apparently localized within the first 300 aa of these proteins. The calmodulin-dependent adenylate cyclase domain of EF, which is homologous to the *B. pertussis* cyclase (Robertson, 1988), occupies its C-terminal 500 aa region. The probable biochemically functional domain for LF, by analogy, probably occupies the corresponding C-terminal region of LF.

(f) Conclusions and perspectives

We have cloned and sequenced the *B. anthracis lef* gene, which encodes LF. LF is part of the tripartite protein exotoxin of *B. anthracis*. The LF-coding gene has features in common with the *B. anthracis* PA-coding (*pag*) and EF-coding (*cya*) genes, including similar RBS and long leader-peptides, which are apparently cleaved during secretion (Robertson et al., 1988; Welkos et al., 1988). Codon usage for each of these toxin genes appeared to be similar, and none of the mature proteins contained Cys. The size of mature LF, deduced from its nt sequence, was 90327 Da, which is close to the experimentally determined values (Leppla et al., 1985; Quinn et al., 1988).

We now know the complete nt sequences for the *B. anthracis cya, lef* and *pag* toxin genes. Using this information, we should be able to construct expression vectors for production of these toxin proteins which can be used for immunological purposes as well as studying the biochemistry of these interesting toxin proteins. In addition, site-specific mutagenesis experiments are in progress to help define regions required for activity, which should lead to the production of a safer recombinant DNA-derived vaccine.

It is also of interest to determine the mechanism by which EF and LF enter the cell. As indicated previously, PA binds to a cell surface receptor and is cleaved prior to binding either EF or LF. Experiments which are designed to determine whether the 300-aa N-terminus of either LF or EF can be used to transport heterologous proteins into a cell using PA as the transporting protein are in progress in our laboratory. The availability of a specific transport mechanism to introduce foreign proteins into mammalian cells could have significant experimental applications.

ACKNOWLEDGEMENTS

This research was supported in part by a contract (DAMD17-85-C-5167) from the U.S. Army Medical Research and Development Command (D.L.R.) and from research funds provided by the Chemistry Department of Brigham Young University. We also express appreciation to Dr. S.H. Leppla for helpful discussions about the *lef* gene and protein.

NOTE ADDED IN PROOF

After this manuscript was accepted, the authors learned that John Lowe (USAMRIID) has also determined the complete *lef* gene sequence with an ORF encoding the same 809-aa LF-precursor protein. J. Lowe used manual sequencing and the Applied Biosystems 370 to determine his sequence.

REFERENCES

Band, L. and Henner, D.J.: *Bacillus subtilis* requires a 'stringent' Shine–Dalgarno region for gene expression. DNA 3 (1984) 17–21.

Beall, F.A., Taylor, M.J. and Thorne, C.B.: Rapid lethal effects in rats of a third component found upon fractionating the toxin of *Bacillus anthracis*. J. Bacteriol. 83 (1982) 1274–1280.

Ezzell, J.W., Ivins, B.E. and Leppla, S.H.: Immunoelectrophoretic analysis, toxicity, and kinetics of in vitro production

of the protective antigen and lethal factor components of *B. anthracis* toxin. Infect. Immun. 45 (1984) 761–767.

Friedlander, A.M.: Macrophages are sensitive to anthrax lethal toxin through an acid-dependent process. J. Biol. Chem. 261 (1986) 7123–7126.

Glaser, P., Ladant, D., Sezer, O., Pichot, F., Ullmann, A. and Danchin, A.: The calmodulin-sensitive adenylate cyclases of *Bordetella pertussis*: cloning and expression in *Escherichia coli*. Mol. Microbiol. 2 (1988) 19–30.

Hopp, T.P. and Woods, K.R.: Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA 78 (1981) 3824–3828.

Ikemura, T.: Correlation between the abundance of *E. coli* transfer RNA and the occurrence of the respective codons in its proteins. J. Mol. Biol. 151 (1981) 389–409.

Kaspar, R.L. and Robertson, D.L.: Purification and physical analysis of *Bacillus anthracis* plasmids pXO1 and pXO2. Biochem. Biophys. Res. Commun. 149 (1987) 362–368.

Leppla, S.H.: Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations in eukaryotic cells. Proc. Natl. Acad. Sci. USA 79 (1982) 3162–3166.

Leppla, S.H.: *Bacillus anthracis* calmodulin-dependent adenylate cyclase: chemical and enzymatic properties and interactions with eukaryotic cells. Adv. Cyclic Nucleotide Protein Phosphorylation Res. 17 (1984) 189–198.

Leppla, S.H., Ivins, B.E. and Ezzell, J.W.: Anthrax toxin. In Leive, L. (Ed.) Microbiology – 1985. American Society for Microbiology, Washington, DC, 1985, pp. 63–66.

Little, S.F. and Knudson, G.B.: Comparative efficacy of *B. anthracis* live spore vaccine and protective antigen vaccine against anthrax in the guinea pig. Infect. Immun. 52 (1986) 509–512.

MacKay, R.M., Lo, A., Willick, F., Zuker, M., Baird, S., Dove, M., Moranelli, F., Seligy, V.: Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nucleic Acids Res. 14 (1986) 9159–9170.

Mead, D.A., Szczesna-Skorupa, E. and Kemper, B.: Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. Protein Eng. 1 (1986) 67–74.

Mock, M., Labruyère, E., Glaser, P., Danchin, A. and Ullmann, A.: Cloning and expression of the calmodulin-sensitive *Bacillus anthracis* adenylate cyclase in *Escherichia coli*. Gene 64 (1988) 277–284.

Moran, C.P., Lang, N., LeGrice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R.: Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. 186 (1982) 339–346.

O'Neill, G.P., Warren, R.A.J., Kilburn, D.G. and Miller Jr., R.C.: Secretion of a *Cellulomonas fimi* exoglucanase by *Escherichia coli*. Gene 44 (1986) 331–336.

Pugsley, A.P. and Schwartz, M.: Export and secretion of proteins by bacteria. FEMS Microbiol. Rev. 32 (1985) 3–38.

Quinn, C.P., Shone, C.C., Turnbull, P.C.B. and Melling, J.: Purification of anthrax-toxin components by high-performance anion-exchange, gel-filtration and hydrophobic-interaction chromatography. Biochem. J. 252 (1988) 753–758.

Robertson, D.L. and Leppla, S.H.: Molecular cloning and expression in *Escherichia coli* of the lethal factor gene of *Bacillus anthracis*. Gene 44 (1986) 71–78.

Robertson, D.L.: Relationships between the calmodulin-dependent adenylate cyclases produced by *B. anthracis* and *B. pertussis*. Biochem. Biophys. Res. Commun. 157 (1988) 1027–1032.

Robertson, D.L., Tippetts, M.T. and Leppla, S.H.: Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): a calmodulin-dependent adenylate cyclase. Gene 73 (1988) 363–371.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Schwarz, W.H., Schimming, S., Rücknagel, K.P., Burgschwaiger, S., Kreil, G. and Staudenbauer, W.L.: Nucleotide sequence of the *celC* gene encoding endoglucanase C of *Clostridium thermocellum*. Gene 63 (1988) 23–30.

Shields, D.C. and P.M. Sharp.: Synonymous codon usage in *B. subtilis* reflects both translational selection and mutational biases. Nucleic Acids Res. 15 (1987) 8023–8040.

Stanley, J.L., Sargent, K. and Smith, H.: Purification of anthrax toxin: factors I and II of the anthrax toxin produced in vivo. J. Gen. Microbiol. 22 (1960) 206–218.

Stanley, J.L. and Smith, H.: Purification of factor I and recognition of a third factor of anthrax toxin. J. Gen. Microbiol. 26 (1961) 49–66.

Thorne, C.B., Molnar, D.M. and Strange, R.E.: Production of toxin in vitro by *Bacillus anthracis* and its separation into two components. J. Bacteriol. 79 (1960) 450–455.

Tinoco Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J.: Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. 246 (1973) 40–41.

Tippetts, M.T. and Robertson, D.L.: Molecular cloning and expression of the *Bacillus anthracis* edema factor toxin gene: a calmodulin-dependent adenylate cyclase. J. Bacteriol. 170 (1988) 2263–2266.

Vodkin, M.H. and Leppla, S.H.: Cloning of the protective antigen gene of *Bacillus anthracis*. Cell 34 (1983) 693–697.

Wang, L.-F., Wong, S.-L., Lee, S.-G., Kalyan, N.K., Hung, P.P., Hilliker, S. and Doi, R.H.: Expression and secretion of human atrial natriuretic α-factor in *Bacillus subtilis* using the subtilisin signal peptide. Gene 69 (1988) 39–47.

Welkos, S.L., Lowe, J.R., Eden-McCutchan, F., Vodkin, M., Leppla, S.H. and Schmidt, J.J.: Sequence and analysis of the DNA encoding protective antigen of *Bacillus anthracis*. Gene 69 (1988) 287–300.

Gene, 81 (1989) 55–64
Elsevier

GENE 03095

Firefly luciferase i
Rhizobium melilo

(Recombinant DNA

Antonio J. Palomare

[a] *Department of Biology (U.S.A.) and* [b] *Departm*

Received by G. Wilcox: 1
Revised: 24 February 198
Accepted: 10 March 1989

SUMMARY

A DNA segment c
pyralis, fused to the λ
82 (1985) 7870–7873
Gram-negative bacter
and extracts prepared
Agrobacterium tumefac
determined by measu
species of *Rhizobium* e
of mature alfalfa nodu
prepared from purifie

INTRODUCTION

The phenomenon o
observed in a wide

Correspondence to: Dr. D.
Genetics, M-034, UCSD
Tel. (619)534-3638; Fax (61
* Present address: Depart
Pharmacy, University of
Tel. 34-54-628355.
** Deceased.

Abbreviations: aa, amino ac
serum albumin; Δ, deletion;